6.0 GUIDANCE TO REGULATORY AUTHORITIES: DETERMINING REASONABLE POTENTIAL AND DERIVING WET PERMIT CONDITIONS

EPA developed the TSD (USEPA 1991a) to support implementation of national policy to control the discharge of toxic pollutants. The TSD presents a statistical approach for determining the need for and the method of deriving water quality-based effluent limits (WQBELs) based on aquatic life (including WET), human health, and wildlife criteria. This approach accounts for the uncertainty associated with small data sets and data variability by assuming a statistical distribution of effluent data (usually lognormal) and calculating a CV or using a default CV to describe data variability.

6.1 Analytical and Sampling Variability in Calculations for Reasonable Potential and Permit Limits

Section 6.1 discusses use of the CV of sample measurements of toxicity to make a reasonable potential determination and to calculate permit limits. Two points must be understood: (1) this CV is to be calculated using toxic unit (TU) values (USEPA 1991a) (see Section 6.2); and (2) EPA strongly recommends that point estimates (not NOEC or LOEC values) be used to calculate the TU values (USEPA 1994a, 1994b).

Water quality-based effluent limits are required when a discharge causes, has reasonable potential to cause, or contributes to an instream excursion above a water quality standard. Throughout this document, EPA uses the commonly understood, shorthand reference "reasonable potential" to refer to this standard for determining the need for a water quality-based effluent limit.

6.1.1 "Adjusting for Analytical Variability" in Calculations for Reasonable Potential and Permit Limits

Adjustment approaches (see Appendix G.3) have been suggested to "adjust for analytical variability" when deriving permit limits and determining the need for a WET limit in the first place. EPA does not recommend these adjustment approaches (Appendix G.3) and strongly reaffirms the statistical approach and methods for calculating permit limits provided in the TSD (USEPA 1991a). *EPA recommends that regulatory authorities use the statistical approach and calculation methods in the TSD*. The TSD methods were designed to provide a reasonable degree of protection for water quality (i.e., to avoid exceedances of water quality criteria), while providing a reasonable degree of protection from the variability of effluent toxicity and analytical variability. The various "adjustment" approaches would undermine these objectives.

The TSD limit calculation for a point source can be divided into two steps: first, convert the wasteload allocation (WLA) to a long-term average (LTA), and then convert the LTA to effluent limits (maximum daily, average weekly, and average monthly limits). WET limit calculations include an intermediate step in which the acute WLA is converted to a WLAa,c. These calculations employ a facility-specific CV based upon effluent sampling data. The TSD approach uses this CV in both steps.

Adjustment approaches intended to account for analytical variability, discussed in detail in Appendix G, would inappropriately use different CVs in these two steps. The first step would use an estimate of the CV of "true" effluent toxicity, which is smaller than the CV for measured toxicities. This approach would result in a larger calculated LTA. The second step would use the CV for the measured toxicities, which is the same CV used in both steps of the TSD approach.

Use of such adjustment approaches would frequently result in setting an average monthly permit limit (AML) that exceeds the chronic WLA. Appendix G demonstrates that such outcomes (i.e., the AML exceeds

the chronic WLA) generally can be expected to occur when various adjustment approaches are used. Appendix G, Table G-1, presents a numerical example of how an adjustment approach would allow calculation of an AML exceeding the chronic WLA (a four-day average value), even when sampling frequency for the calculation is set at the recommended minimum of four samples per month. [It is acceptable for the maximum daily limit (MDL), which applies to a single sample, to exceed the chronic WLA. It is also acceptable for the AML to exceed the chronic WLA, if the AML calculation is based on fewer than four samples per month. Note, however, that the TSD recommends always assuming at least four samples per month when calculating the AML.]

The TSD reasonable potential calculation first calculates the percentile represented by the maximum observed TU value. For example, the maximum of 10 reported TU values is identified with the 63rd percentile. Then the sample CV is used to project the 95th or 99th percentile TU value, using a table of reasonable potential multiplying factors. This value is combined with the appropriate mixing-zone dilution to project a maximum receiving water toxicity, which is compared with the applicable water-quality criterion. If an adjustment were applied to the reasonable potential calculation, the CV would be adjusted downward and the maximum projected receiving water toxicity would be smaller. This would make a determination of need for a permit limit less likely.

Because of these considerations, EPA strongly recommends that no adjustment be made to the CV or variance of toxicity, either for reasonable potential or permit limit calculations. The TSD statistical approaches already account for analytical variability appropriately. EPA continues to recommend the TSD approach, which ensures that effluent limits and, thereby, *measured* effluent toxicity or pollutant parameter concentrations are consistent with calculated WLAs.

6.1.2   Analytical Variability and Self-monitoring Data

EPA determines compliance with permit limits on the basis of self-monitoring data, and these data include some measure of analytical variability. The influence of analytical variability is accounted for in the TSD statistical procedures used to set water-quality limits and determine the potential for toxicity, as explained in Appendix G.

The permittee is responsible for ensuring that measured discharge toxicity never exceeds the permit limits. No special allowance is made for analytical variability in assessing compliance. The maximum discharge toxicity should incorporate a margin of safety, which will account for sampling and analytical variability. In other words, to avoid exceeding permit limits, the facility's treatment system should be designed so that the maximum toxicity is somewhat lower than its permit limits.

6.1.3   Precision of WET Measurements and Estimates of Effluent CV

Single measurements on effluent involve some uncertainties about the true concentration or toxicity related to representativeness of the sample, including sample holding time and conditions, and the analytical measurement system. Like all analytical measurements, WET measurements (NOEC, EC25, LC50) are inexact. That is, the exact toxicity of an analyte in a sample can be specified only within some range. This imprecision can be reduced by using a suitable number of organisms and replicates for each test (see Section 5.3.2 on experimental design).

The numbers of organisms and replicates required for EPA WET method test acceptability are specified as minimums. Test precision will be approximately proportional to the square root of the number of replicates. Thus, doubling the number of replicates may decrease the MSD to approximately 70 percent of its former value. Increased replication also tightens the confidence interval for a point estimate of the effect concentration (e.g., EC25 and LC50).

EPA strongly recommends that toxicity measurements of an effluent be obtained at least quarterly for three years to provide a good basis for determining the need for limits and for calculating limits. One year should be regarded as the minimum duration needed to characterize effluent variability (due to seasonal, stream flow, or process fluctuations), and ten the minimum number of measurements, unless scientific and technical knowledge supports a shorter period as representative of the distribution of pollutant types and concentrations of toxicity.

Estimates based on multiple measurements involve the same uncertainties that apply to single measurements. They also may involve larger uncertainties related to sampling error, that is, the chance that typical levels of toxicity or concentrations of pollutant may not be encountered during the sampling program. The sampling program may not fully characterize effluent variability if too few samples are taken, the sampling times and dates are not representative, or the duration of the sampling program is not long enough to represent the full range of effluent variability. When determining the need for limits and calculating limits, the variance or the CV of toxicity measurements is key. The larger the number of samples, the more precise is the estimate. Confidence intervals for the variance and CV can be calculated and carried through the calculations for reasonable potential and effluent limits (Appendix G). Even when assumptions are not strictly met, confidence intervals provide a useful perspective on the uncertainty of the results and the need for more samples. The *minimum* number of measurements recommended for calculating estimates of the CV for effluent toxicity is 10.

6.1.4   Between-Laboratory Variability

Between-laboratory variability may increase the CV as discussed in Section 6.1.1, if the toxicity tests were conducted by more than one laboratory for a specific facility. A concern to permittees is that this may increase the likelihood of making a finding of reasonable potential.

Within-laboratory variability is the component of analytical variability that should be reflected in regulatory calculations. If the data used for reasonable potential or permit limit calculations are effluent measurement data reported by at least two laboratories, there are ways to appropriately estimate the variance to be used in TSD statistical calculations.

For example:

- If the same laboratories continue to be used in the same proportion or frequency and the measurements from the individual laboratories represent different sampling dates, the measurement data can be treated as if they were generated by a single laboratory. This approach may increase the estimated variance and the AML, which is not in the interest of the permittee. Selecting one laboratory for future monitoring, based on the variance of its reported reference toxicant test results, should mitigate this problem.

- If only one laboratory has reported data on each sampling date, and the other laboratories report over different time spans or over the same time span on alternating dates, EPA recommends forming a pooled estimate of variance. Calculate the sample variance (S^2) of log(TU) for each laboratory separately, and combine these using the formula:

    pooled variance of $\log(X) = [(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2] / [(N_1 - 1) + (N_2 - 1)]$

    An analogous formula is used for more than two laboratories. The same result can be obtained by conducting a one-way analysis of variance on log(TU) (with laboratories treated as the groups or classes) and using the reported EMS.

Changing a laboratory may change analytical (within-laboratory) variability of measurements and test sensitivity (i.e., PMSD values). That is, the average effect concentration may change (e.g., Warren-Hicks et al. 1999). Ideally, the permittee will anticipate and plan for a change of testing laboratory. Permittees should compare reference toxicant test data for current and candidate replacement laboratories, selecting one with acceptable variability and a similar average effect concentration.

6.2 Determining Reasonable Potential and Establishing Effluent Limits

Effluent characterization is an essential step in determining the need for an NPDES permit limit. NPDES regulations under 40 CFR Part 122.44(d)(1)(ii) specify that reasonable potential include "*whether a discharge causes, has the reasonable potential to cause, or contributes to an instream excursion above a State water quality standard*." Calculations for reasonable potential determination and for permit limits should follow EPA guidance in the TSD (USEPA 1991a). In particular, the TSD statistical methods should be used. Such calculations should use TUs for WET data, not effect concentrations (percent whole effluent). Toxic units are defined (USEPA 1991a, Chapter 1.3.1, page 6) as the reciprocal of the effect concentration times 100, where the effect concentration is expressed as a percentage of whole effluent, thus $TUa = 100/LC50$ and $TUc = 100/ECp$.

When characterizing an effluent to determine whether a permit limit is necessary, permit writers can use the available effluent WET data and a water-quality model to perform a reasonable potential analysis. The TSD outlines the statistical approach. This approach uses existing effluent data to project a maximum pollutant concentration or a maximum toxicity in the effluent (USEPA 1991a). The projected maximum concentration or toxicity is used as an input in the water quality model to determine whether the effluent has the reasonable potential to cause or contribute to an excursion of ambient water quality criteria. If reasonable potential exists, the permit writer must derive a WET permit limit for that facility.[1]

The variability of the existing effluent data, as measured by the CV, has a significant effect on the projected maximum pollutant concentration or toxicity. The higher the CV, the higher the projected maximum, and the more likely that there is reasonable potential and a limit is needed. EPA recommends that regulatory authorities use all valid, relevant, and representative data in making reasonable potential determinations. EPA is developing a national policy clarifying use of the TSD procedures for determining reasonable potential for WET. Important components of this policy include specifying the minimum number of valid WET tests necessary to calculate facility-specific CVs,[2] as well as recommending a step-wise approach to determining the need for WET permit limits. This approach reflects a strong preference by EPA and its stakeholders to rely on facility-specific WET testing, based on adequate frequency and duration of effluent sampling, for making reasonable potential determinations for toxicity.

EPA recommends that point estimates be used to estimate effluent variability, to determine the need for limits, and to set permit limits. This is recommended whether the self-monitoring test results will be determined using hypothesis tests or point estimates. Point estimates have less analytical variability than NOECs using current experimental designs, as shown in Chapter 3. Point estimates make the best use of the WET test data for purposes of estimating the CV, LTA, and RP factor and calculating the permit limit.

---

[1]   When the State has narrative criteria for toxicity and the TIE/TRE identifies a specific chemical that is the source of toxicity, the permit writer may include a chemical-specific limit for that parameter instead of a WET permit limit in accordance with 40 CFR Part 122.44(d)(v).

[2]   If fewer than ten data points are available, the regulatory authority must use a default CV. As a result, the need for a WET permit limit may be based on a default value rather than actual data.

6.3 Development of a Total Maximum Daily Load for WET

Total maximum daily loads (TMDLs) may be indicated when there is acute or chronic toxicity in a waterbody, leading to the listing of the waterbody as impaired under CWA Section 303(d), and when there are multiple sources of the toxicity. EPA believes that TMDL calculations should be performed on the pollutants causing toxicity whenever possible. In these situations, EPA suggests that the first step of the analysis is to conduct ambient toxicity identification evaluations to identify the pollutant(s) and the source(s) causing the toxicity. Once the pollutant(s) and source(s) causing toxicity have been identified for the waterbody, then a TMDL should be developed for the individual pollutant(s).

6.4 Accounting for and Minimizing Variability In the Regulatory Decision Process

A common goal for the permittee and the regulatory authority is to have confidence in the test results from the biological and statistical procedures. Both permittees and regulatory authorities would then have more confidence in taking regulatory actions, such as evaluating multiple effluent samples to determine reasonable potential and derive permit conditions (e.g., permit limits, monitoring triggers). If steps such as collecting a representative effluent sample to conducting the toxicity tests properly, as discussed in Sections 5.2 through 5.4, and requiring additional TACs (Section 6.4.1) are used to reduce or minimize within-test variability, then the reliability of the WET test results increases.

6.4.1   Recommended Additional TACs: Lower and Upper Bounds for PMSD

Reference toxicant data from a large number of tests and laboratories were used to generate PMSD values; percentiles of these values are reported in Table 3-6. The MSD represents the smallest difference between the control mean and a treatment mean that leads to the statistical rejection of the null hypothesis (i.e., no toxicity) using Dunnett's multiple comparison test. MSD values are divided by the control mean and multiplied by 100 to produce a "percent MSD" (PMSD) value. The PMSD allows comparison of different tests and represents the smallest significant difference from the control as a percentage of the control mean. Thus, it represents the smallest significant value of the relative difference [100 (control mean - treatment mean)/control mean]. The MSD is often expressed as a percentage of the biological endpoint in the control response.

The following formula is used to calculate MSD (as recommended by USEPA 1995):

$$MSD = d\,s_w\sqrt{(1/n_1)+(1/n)}$$

where

d = critical value for the Dunnett's procedure
s_w = the square root of the error mean square (EMS)
n_1 = number of experimental units in the control treatment
n = the number of experimental units per treatment, assuming an equal number at all other treatments

Percent MSD is calculated as follows:

$$PMSD = \frac{MSD}{control\ mean} \times 100$$

EPA recommends that regulatory authorities implement both the lower and upper PMSD bound approach to minimize within-test variability when using hypothesis testing approaches to report an NOEC. The implementation of the upper PMSD bound should also apply when using point estimate techniques. There are five possible outcomes for regulatory decisions (see Figure 6-1). Two outcomes imply unqualified acceptance of the WET test statistical result:

1. **Unqualified Pass**–The test's PMSD is within bounds and there is no significant difference between the means for the control and the instream waste concentration (IWC) treatment. The regulatory authority would conclude that there *is no toxicity at the IWC concentration*.

2. **Unqualified Fail**–The test's PMSD is larger than the lower bound (but not greater than the upper bound) in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The regulatory authority would conclude that there *is toxicity at the IWC concentration*.

3. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is no significant difference between the means for the control and the IWC treatment. The test is considered invalid. A new effluent sample must be collected and another toxicity test must be conducted.

4. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The test is considered valid. The regulatory authority would conclude that there *is toxicity at the IWC concentration.*

5. **Very Small but Significant Difference**–The relative difference (see Section 6.4.2, below) between the means for the control and the IWC treatment is smaller than the lower bound in Table 3-6 and this difference is statistically significant. The test is acceptable. The NOEC is determined as described in Section 6.4.2 below.
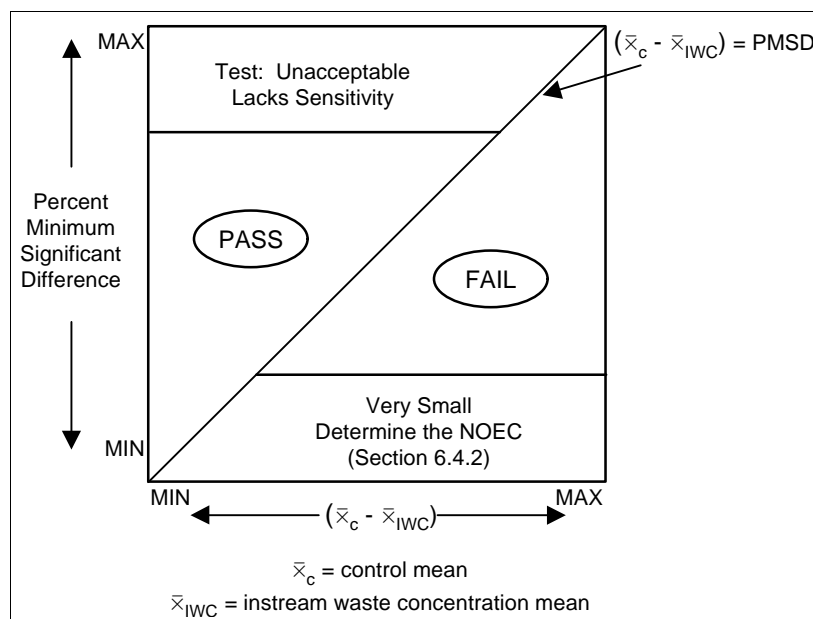

Figure 6-1. Paradigm that incorporates the lower and upper percent minimum significant difference.

Regulatory authorities should examine the sample permit language as provided in Appendix C, for incorporation of the PMSD bound language in a NPDES permit.

Note that "unqualified acceptance" of a WET test result requires that all of the following must be achieved: (1) collect the effluent sample properly; (2) conduct the toxicity test methods as specified in the toxicity manuals; (3) meet the required TACs; (4) meet the proper water quality parameters (e.g., temperature, pH); and (5) conduct the proper statistical calculations. All these conditions must be reviewed and deemed acceptable before a test is evaluated for self-monitoring data and reporting.

Figure 6-2 provides a decision tree that regulatory authorities can use when implementing the lower and upper PMSD bounds.
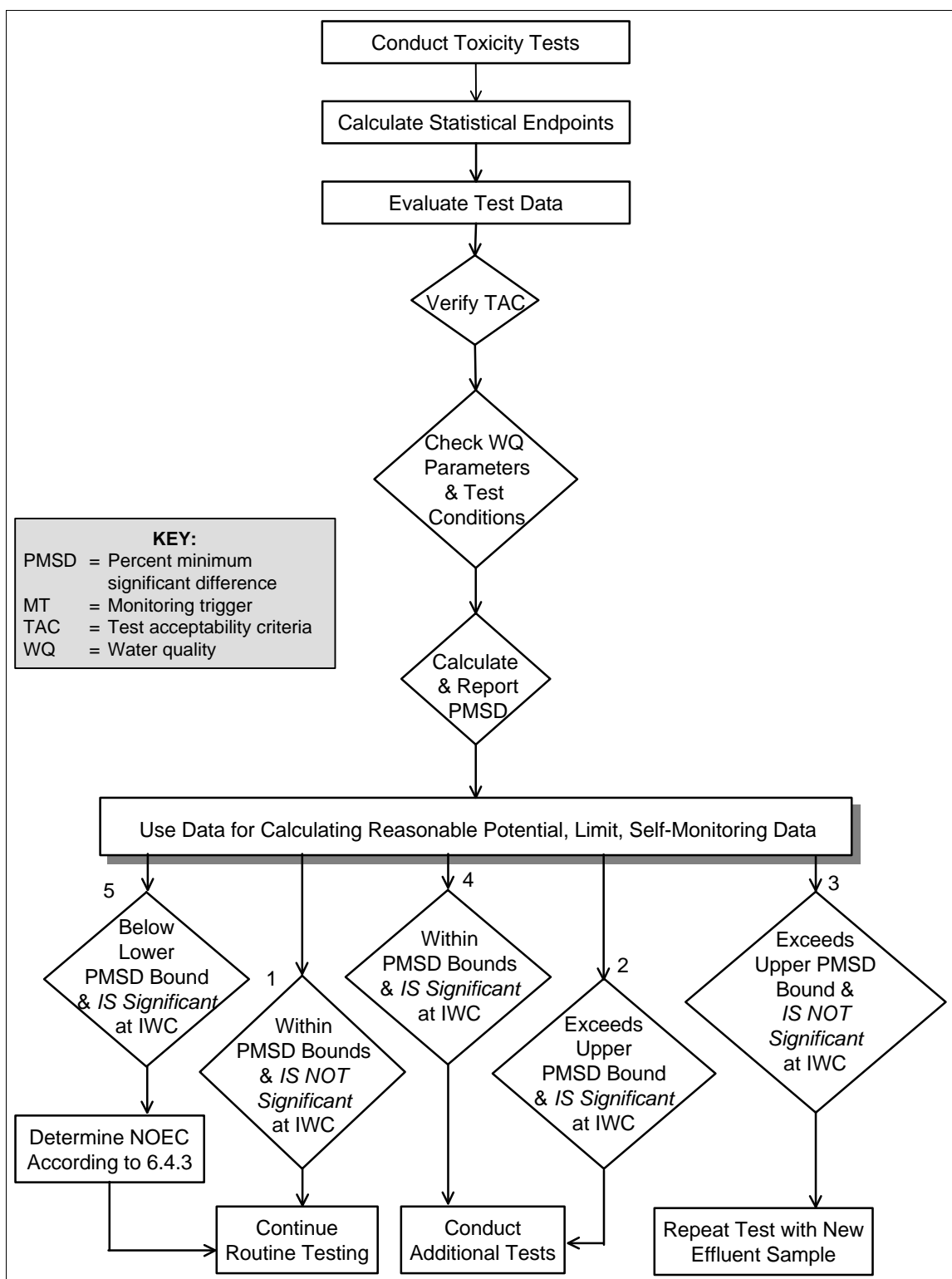
Figure 6-2. Implementing applications of upper and lower PMSD bounds for effluent toxicity testing requirements.

6.4.2   How to Determine the NOEC Using the Lower PMSD Bound

If the permit specifies that self-monitoring data are to be generated using hypothesis testing approaches, then the analyst should report the NOEC as the following. Find the smallest concentration for which (a) the treatment mean differs significantly from the control mean and (b) the relative difference (see example below) is not smaller than the 10th percentile in Table 3-6. Therefore, the NOEC is the next smaller test concentration.

In other words, concentrations having a very small relative difference with control (smaller than the lower PMSD bound) would be treated as if they do not differ significantly from control (even if they do so), for the purpose of determining the NOEC.

Table 6-1 illustrates the application of the lower PMSD bound for the reproduction endpoint of a *Ceriodaphnia* chronic test. In this example, the test's PMSD was 9.9, smaller than the 10th percentile value of 11 found in Table 3-6. The IWC concentration differed significantly from the control. The test falls under outcome number 5, a significant but very small difference at the IWC. The first step is to calculate the relative differences from control (Table 6-1) as [(control mean - treatment mean) divided by (control mean)] × 100. The next step is to determine which relative differences exceed the PMSD lower bound, 11 in this case (see the last column of Table 6-1). Finally, the NOEC is determined as described above. The NOEC is 12.5 percent effluent for this example.

Table 6-1.   Example of Applying the Lower PMSD Bound for the Chronic
Ceriodaphnia Test with the Reproduction Endpoint

Concentration (percent effluent)	Reproduction (mean of ten replicates)	Relative Difference from Control	Does Relative Difference Exceed 11?
100%	5.08 *	82	Yes
50%	12.4 *	56	Yes
25%	23.4 *	17	Yes
IWC = 12.5%	25.3 *	10	No
6.25%	26.1	7.4	No
Control	28.2	0	No

NOTE: The lower PMSD bound for this method and endpoint is 11 (Table 3-6). In this example, the NOEC is 6.25 percent effluent using the test's (very small) PMSD. Therefore, the reported NOEC should be 12.5 percent effluent after applying the lower PMSD bound.

*   Differs statistically from the control as determined by MSD = 2.8 neonates. Thus, treatment means that are less than 28.2 - 2.8 = 25.4 would be statistically significant. These correspond to relative differences greater than 100 (2.8 / 28.2) = 9.9 percent.

6.4.3   Justification for Implementing the Test Sensitivity Bounds

A lower bound is needed to avoid penalizing laboratories that achieve unusually high precision. The 10th percentile PMSD represents a practical limit to the sensitivity of the test method because few laboratories are able to achieve such precision on a regular basis and most do not achieve it even occasionally. Several independent researchers have evaluated and provide support for using the MSD approach as additional TAC for the toxicity test methods. Thursby et al. (1997) advocate and provide reasons for using an empirical data base of minimum significant differences to provide TAC using statistical performance assessment. The State of California (Hunt et al. 1996, Starrett et al. 1993) and the West Coast marine toxicity test methods (USEPA 1995) have implemented an upper PMSD bound to minimize insensitive tests. Also the State of North Carolina has implemented additional requirements for the *Ceriodaphnia* chronic tests that reduced method

variability. North Carolina's evaluation of these additional TACs and subsequent improvements in test sensitivity appears in Appendix F.

The North Carolina data base affords the opportunity to evaluate the effectiveness of additional TAC and changes to the toxicity test procedures as they relate to the variability of WET test results (see Appendix F). For example, for PMSD, the median value decreased from 21 percent to 16 percent, while the 90th percentile decreased from 39 percent to 31 percent, indicating an overall increase in test sensitivity. The range in median values across all laboratories before adopting additional TACs was 12 percent to 36 percent. After adopting additional TACs, the range in median values was 10 percent to 27 percent, indicating a decrease in the overall spread between laboratories. The range in control CVs within a laboratory was from 21 percent to 79 percent before adopting TACs, compared to the range in control CVs within a laboratory after adopting TACs, which was narrowed to 17 percent to 36 percent. Overall, laboratories are generating data with more consistency within and between laboratories, after implementation of the additional TACs and additional method guidance provided by the State for the chronic *Ceriodaphnia dubia* test method.

6.4.4 Guidance to Testing Laboratories on How to Achieve the Range of Performance for PMSD

EPA recommends that regulatory authorities use the upper bounds (90th percentiles for PMSD in Table 3-6) to identify tests that are insufficiently sensitive. If PMSD exceeds this upper bound more often than occasionally, the laboratory should thoroughly investigate ways to reduce variability. There are three principal ways to reduce PMSD: (1) decrease within-test variability (that is, decrease the error mean square and therefore the standard deviation at each concentration); (2) increase the control mean; and (3) increase the number of replicates. The number of replicates required could be determined by trial-and-error calculations using the error mean square values obtained from a series of WET tests. At least 20 tests are recommended. The number "n" in the formula for MSD (number of replicates) would be increased and MSD re-calculated for each error mean square value. This approach uses a sample of tests specific to a particular laboratory and reveals the variation among tests. This approach would demonstrate how many replicates would be needed to achieve the upper PMSD bound, as required in Table 3-6.

6.5 Additional Guidance That Regulatory Authorities Should Implement to Further Support the WET Program

As discussed in Section 5.3, regulatory authorities have the discretion to develop and implement additional WET program requirements and guidance to ensure that WET test method variability is reduced by specifying additional guidance beyond the minimum requirements of EPA's WET test method's QA/QC and TACs. Appendix E provides a snapshot of State approaches to implementing NPDES WET programs to minimize WET test variability.

These State approaches include WET information to assist the regulated community with the following:

- Guidance regarding the evaluation of reference toxicant and effluent test results

- Guidance regarding how the State reviews reference toxicant data for laboratory performance

- Guidance regarding additional QA/QC criteria the State has developed and implemented

- Guidance regarding efforts the State has made to minimize test method variability

- Description of how the State reviews or conducts performance laboratory audits

- Description of specific implementation guidance that the State has developed to assist permit writers

- Description of how the State provides or uses toxicity test training

States contemplating such changes should consult with EPA to ensure the changes will be appropriate in the context of the State's overall NPDES WET program. In addition, States should implement a step-wise approach to address toxicity when the permit limit or monitoring trigger is exceeded in their State WET implementation plans.

For example, when an effluent is deemed toxic, then the permittee should take appropriate steps to demonstrate the magnitude, frequency, and potential source(s) of the toxicity. The components of the step-wise approach could include increased frequency of toxicity testing to characterize the magnitude and frequency of toxicity. If continued toxicity is demonstrated, then the permittee could conduct a Toxicity Reduction Evaluation/Toxicity Identification Evaluation (TRE/TIE) with toxic effluent sample(s) (USEPA 1991b, 1992). For example, EPA Regions 9 and 10 have prepared WET implementation guidance to assist their States (Denton and Narvaez 1996). This guidance provides sample permit language for a step-wise approach to address toxic samples (see Appendix C).

6.6 Chapter Conclusions

The TSD statistical approach to reasonable potential determination and permit limit derivation considers combined effluent and analytical variability through the CV of measured effluent values. Because determination of effluent variability is based on empirical measurements, the variability estimated for effluent measurements includes the variability of pollutant levels, sampling variability, and a smaller component owed to method variability. Steps should be taken to reduce these sources of variability. EPA believes that the TSD statistical procedures are appropriately protective in considering both effluent and analytical variability in reasonable potential and effluent limit calculations.

EPA recommends that regulatory authorities use a sampling program that conducts at least ten representative WET tests over a period of three years to represent the full range of effluent variability. Regulatory authorities should use recommended procedures in the TSD to determine when numeric WET limits or WET monitoring triggers are needed. Other permit conditions may include monitoring triggers, such as increased toxicity testing, TREs/TIEs, and follow-up actions initiated because a permit limit is exceeded or a monitoring trigger is not met. Regulatory authorities should implement the additional test sensitivity requirements by requiring that each test result not exceed the upper PMSD bound. In addition, regulatory authorities should determine the appropriate NOEC for test results below the lower PMSD bound as described in Section 6.4.2. These efforts should lead to increased confidence in the effect concentrations that are generated to evaluate self-monitoring data.